BIDS

BRAIN IMAGING DATA STRUCTURE

(for EEG)

Cyril Pernet, PhD
Neurobiology Research Unit,
Copenhagen University Hospital, Rigshospitalet

30th EEGLAB workshop – 14th June 2021
https://doi.org/10.6084/m9.figshare.14748105

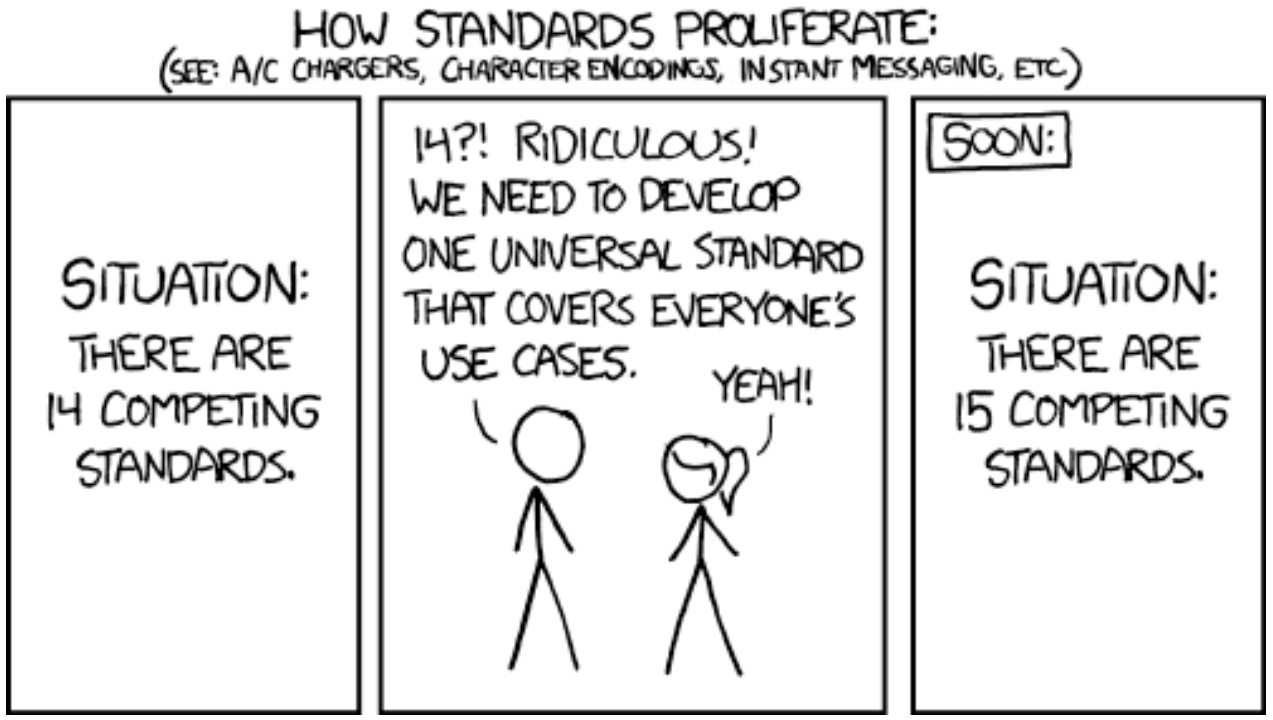
@CyrilRPernet
cyril.pernet@nru.dk

What is BIDS?

- It’s a data **structure** ; nothing to do with **format** per se
- It’s about: - how you organize data in a folder
  - how you name files
  - how you document metadata
  - using community standards and dictionaries to do all the above
- It cares about imaging data but also behaviour/cognition

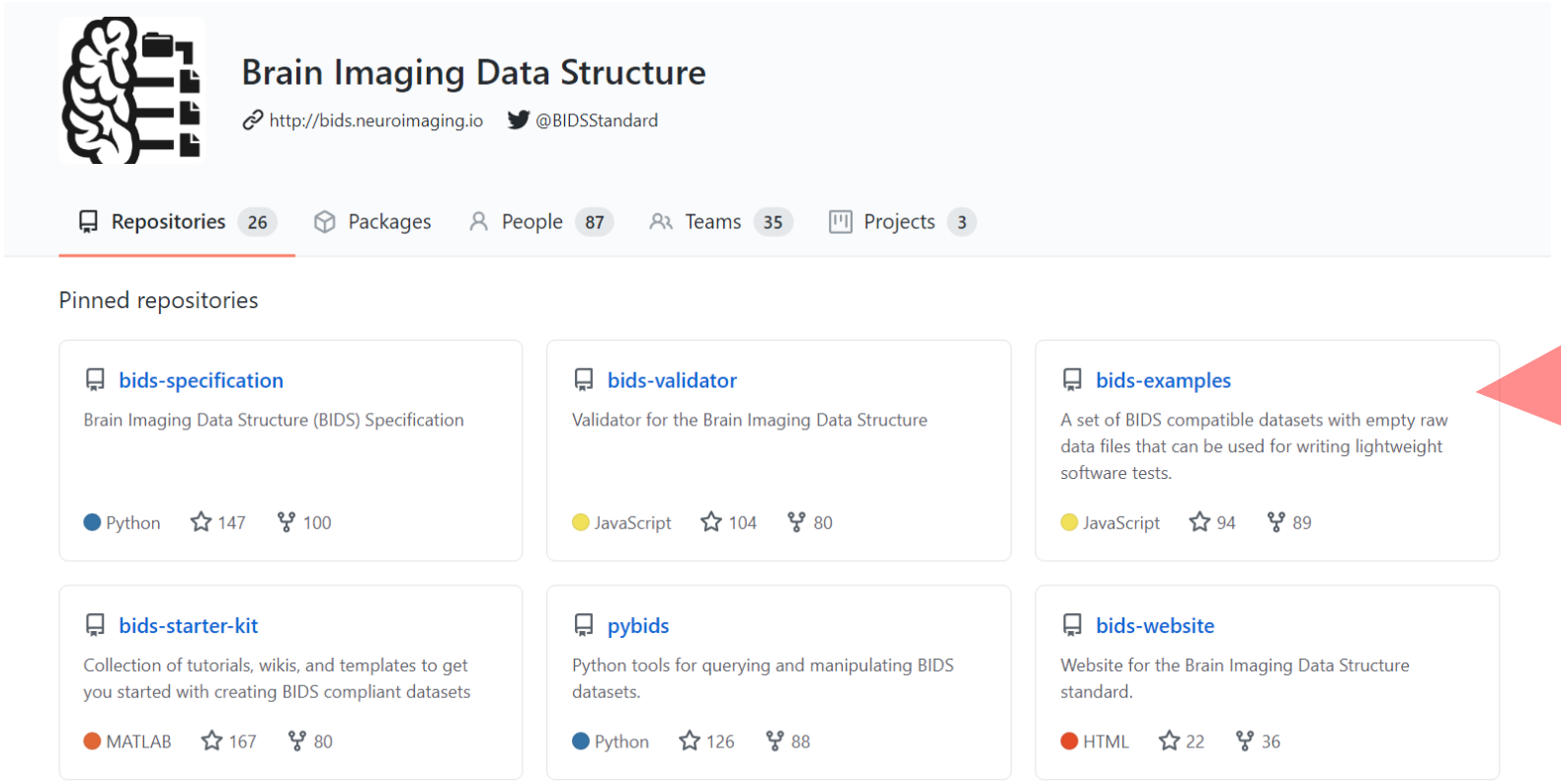
bids.neuroimaging.io

BIDS is a widely supported “best practice” (becoming *a posteriori*, the “standard”, rather than been created *de novo*)



Oostenveld & Zwiers – 2020 BIDS talk

BIDS is a widely supported “best practice”

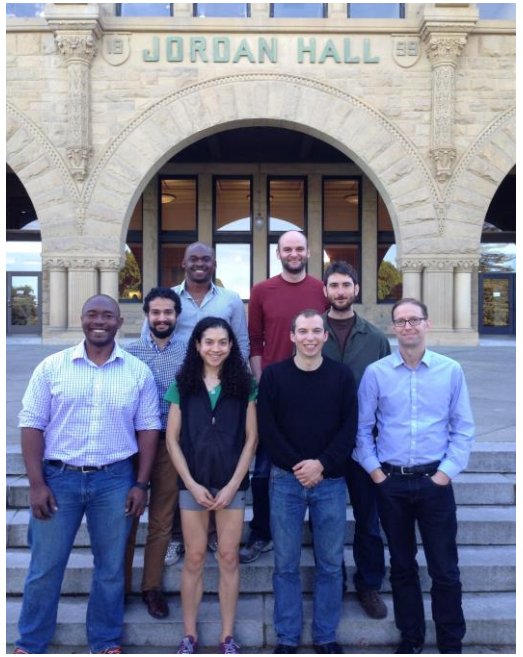


Everyone is welcome to contribute – examples, tests, documentation, code

https://github.com/bids-standard

CREDITS

Data Sharing Task Force



The Poldrack Lab @ Stanford

SCIENTIFIC DATA

OPEN

SUBJECT CATEGORIES

» Data publication and archiving

» Research data

The brain imaging data structure, a format for organizing and describing outputs of neuroimaging experiments

Krzysztof J. Gorgolewski[1], Tibor Auer[2], Vince D. Calhoun[3,4], R. Cameron Craddock[5,6], Samir Das[7], Eugene P. Duff[8], Guillaume Flandin[9], Satrajit S. Ghosh[10,11], Tristan Glatard[7,12], Yaroslav O. Halchenko[13], Daniel A. Handwerker[14], Michael Hanke[15,16], David Keator[17], Xiangrui Li[18], Zachary Michael[19], Camille Maumet[20], B. Nolan Nichols[21,22], Thomas E. Nichols[20,23], John Pellman[6], Jean-Baptiste Poline[24], Ariel Rokem[25], Gunnar Schaefer[1,26], Vanessa Sochat[27], William Triplett[1], Jessica A. Turner[3,28], Gaël Varoquaux[29] & Russell A. Poldrack[1]

Received: 18 December 2015
Accepted: 19 May 2016
Published: 21 June 2016

Evolution of BIDS

1. Kickoff meeting at Stanford in Spring 2015
2. Meeting at OHBM 2015 (June)
3. Introduced to neuroinformatics community at INCF Congress 2015 (August)
4. First release candidate and public call for comments (September)
5. Version 1.0.0 published along the introductory paper

→ Initially covered structural MRI and fMRI, now all sort of MRI, PET, EEG, MEG, iEEG, extensions for animal, connectivity, imaging genomics ..

Datatype specific publications

EEG

- Pernet, C. R., Appelhoff, S., Gorgolewski, K.J., Flandin, G., Phillips, C., Delorme, A., Oostenveld, R. (2019). **EEG-BIDS, an extension to the brain imaging data structure for electroencephalography**. Scientific data, 6 (103). doi:10.1038/s41597-019-0104-8

iEEG

- Holdgraf, C., Appelhoff, S., Bickel, S., Bouchard, K., D'Ambrosio, S., David, O., Devinsky, O., Dichter, B., Flinker, A., Foster, B. L., Gorgolewski, K. J., Groen, I., Groppe, D., Gunduz, A., Hamilton, L., Honey, C. J., Jas, M., Knight, R., Lauchaux, J.-P., Lau, J. C., Lee-Messer, C., Lundstrom, B. N., Miller, K. J., Ojemann, J. G., Oostenveld, R., Petridou, N., Piantoni, G., Pigorini, A., Pouratian, N., Ramsey, N. F., Stolk, A., Swann, N. C., Tadel, F., Voytek, B., Wandell, B. A., Winawer, J., Whitaker, K., Zehl, L., Hermes, D. (2019). **iEEG-BIDS, extending the Brain Imaging Data Structure specification to human intracranial electrophysiology**. Scientific data, 6 (102). doi:10.1038/s41597-019-0105-7

MEG

- Niso Galan, J.G., Gorgolewski, K.J., Bock, E., Brooks, T.L., Flandin, G., Gramfort, A., Henson, R.N., Jas, M., Litvak, V., Moreau, J., Oostenveld, R., Schoffelen, J.-M., Tadel, F., Wexler, J., Baillet, S. (2018). **MEG-BIDS, the brain imaging data structure extended to magnetoencephalography**. Scientific Data, 5 (180110). doi:10.1038/sdata.2018.110

The BIDS goal is to make more data accessible to more researchers

Making more data accessible

- for yourself in 6 months time

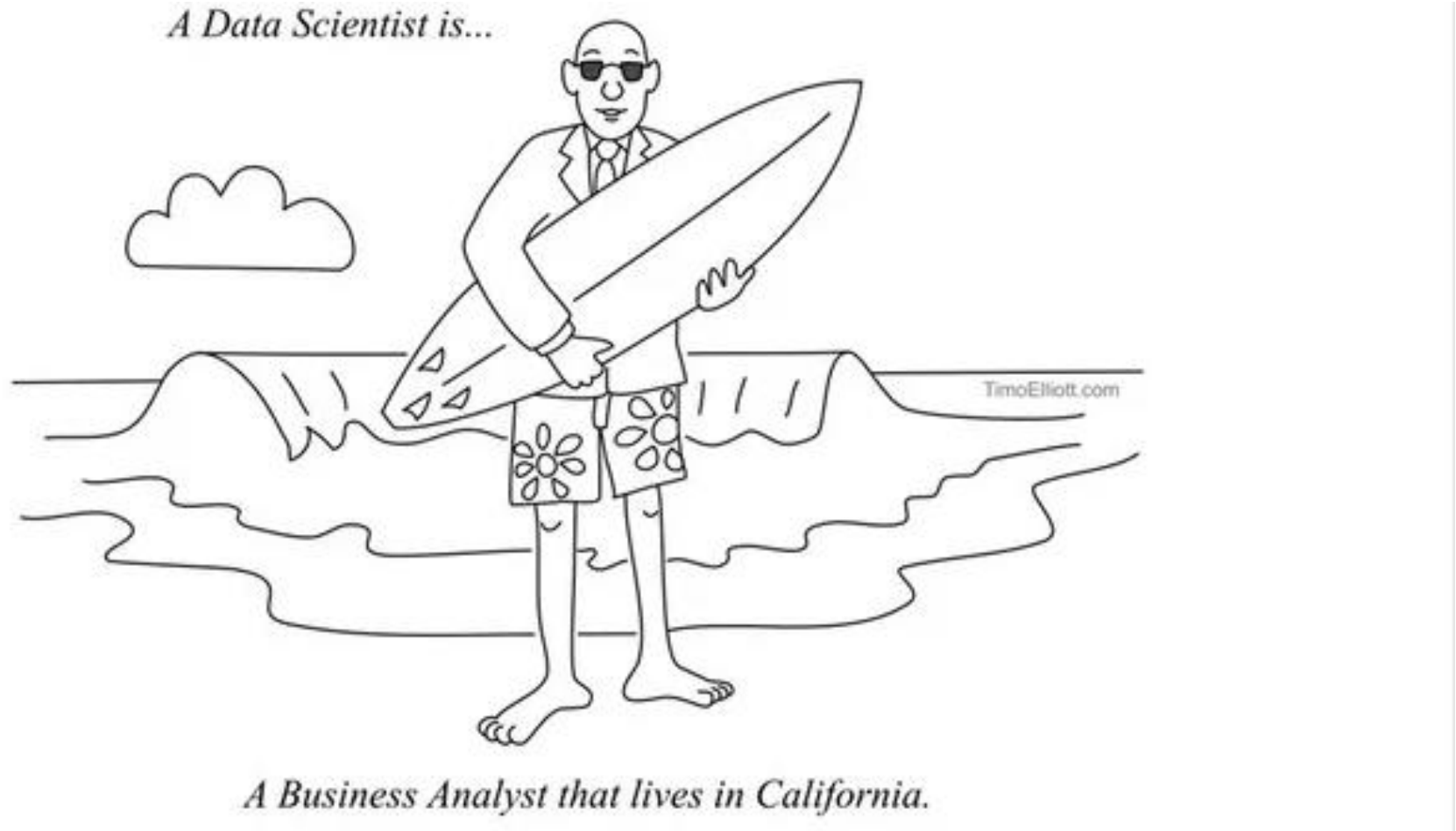
- to people in the lab
(new students, collaborators, governance)

- to other researchers
(data sharing)

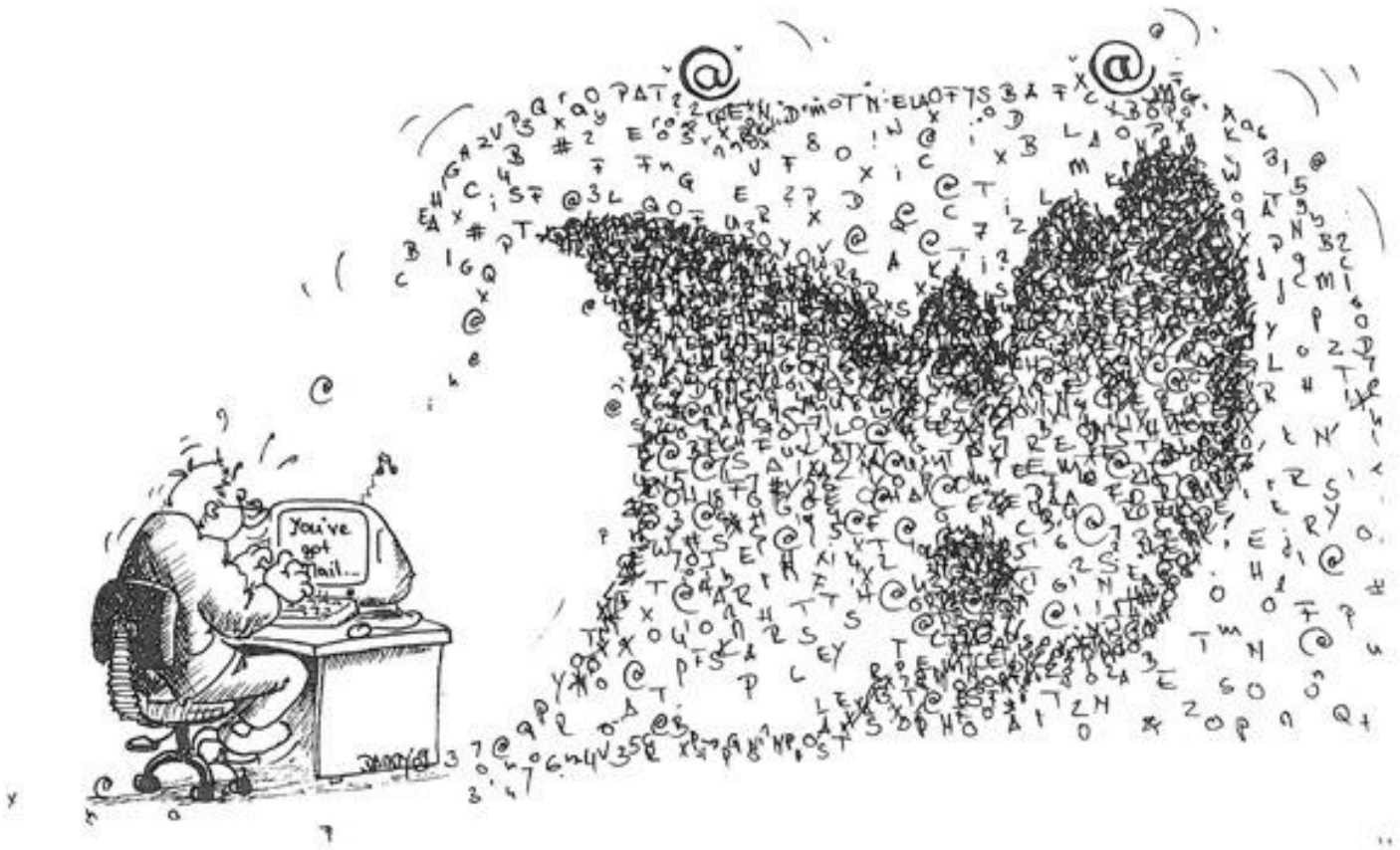
Meet Prof. Smith

I need to ask Mike about these ‘old’ data, there is so much more we can do with this

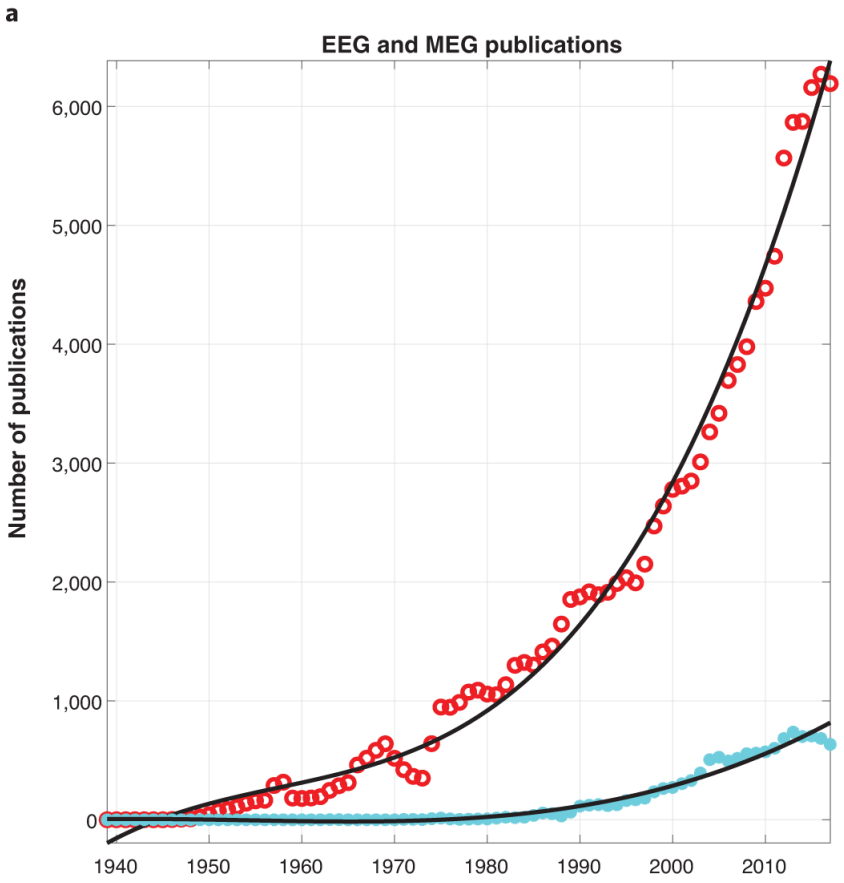
Meet Mike


A Data Scientist is...
TimoElliott.com
A Business Analyst that lives in California.

Prof Smith is lost in her data

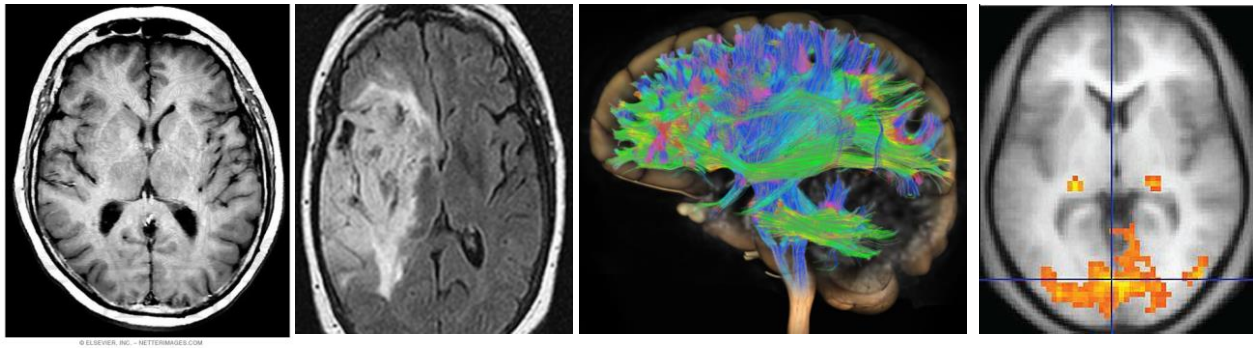


Getting lost in your data



COBIDAS MEEG – figure 1

- **Multitude of techniques** (MEG, EEG, PET, MRI, NIRS, TMS, etc …) and applications.
- Despite similarities in experimental designs and data types **each researcher tends to organize and describe their data in their own way**.



BIDS principles

Principles

1. **Adoption is crucial**

= got to fit peoples' needs = driven by user cases

2. Don't reinvent the wheel

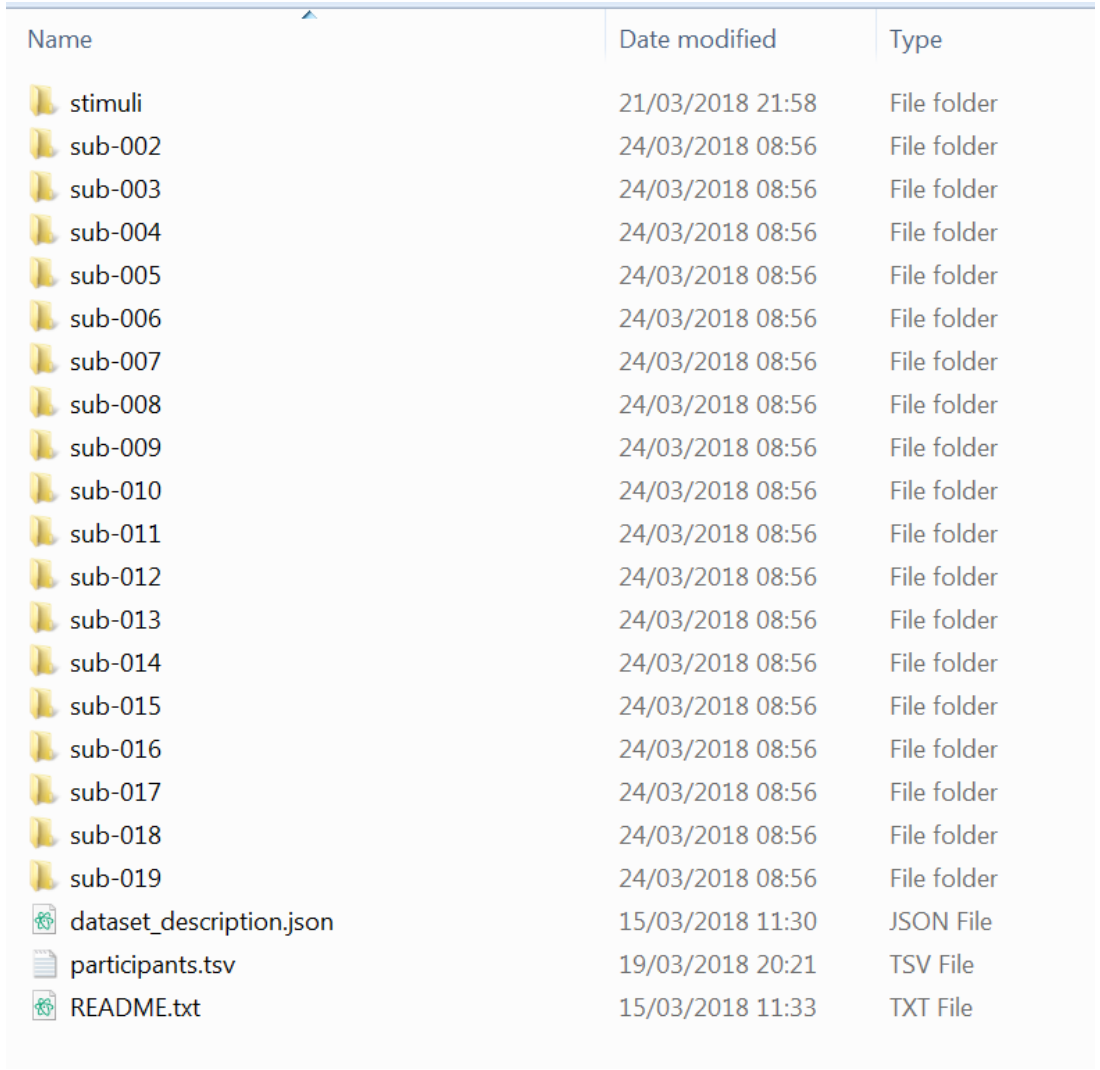
= use already in use file formats = minimize work for users

3. 80/20 rule

= 80% of the work is already done by you collecting data, just need 20% to get it to BIDS

BIDS data structure basics

A BIDS folder (any/all modality)

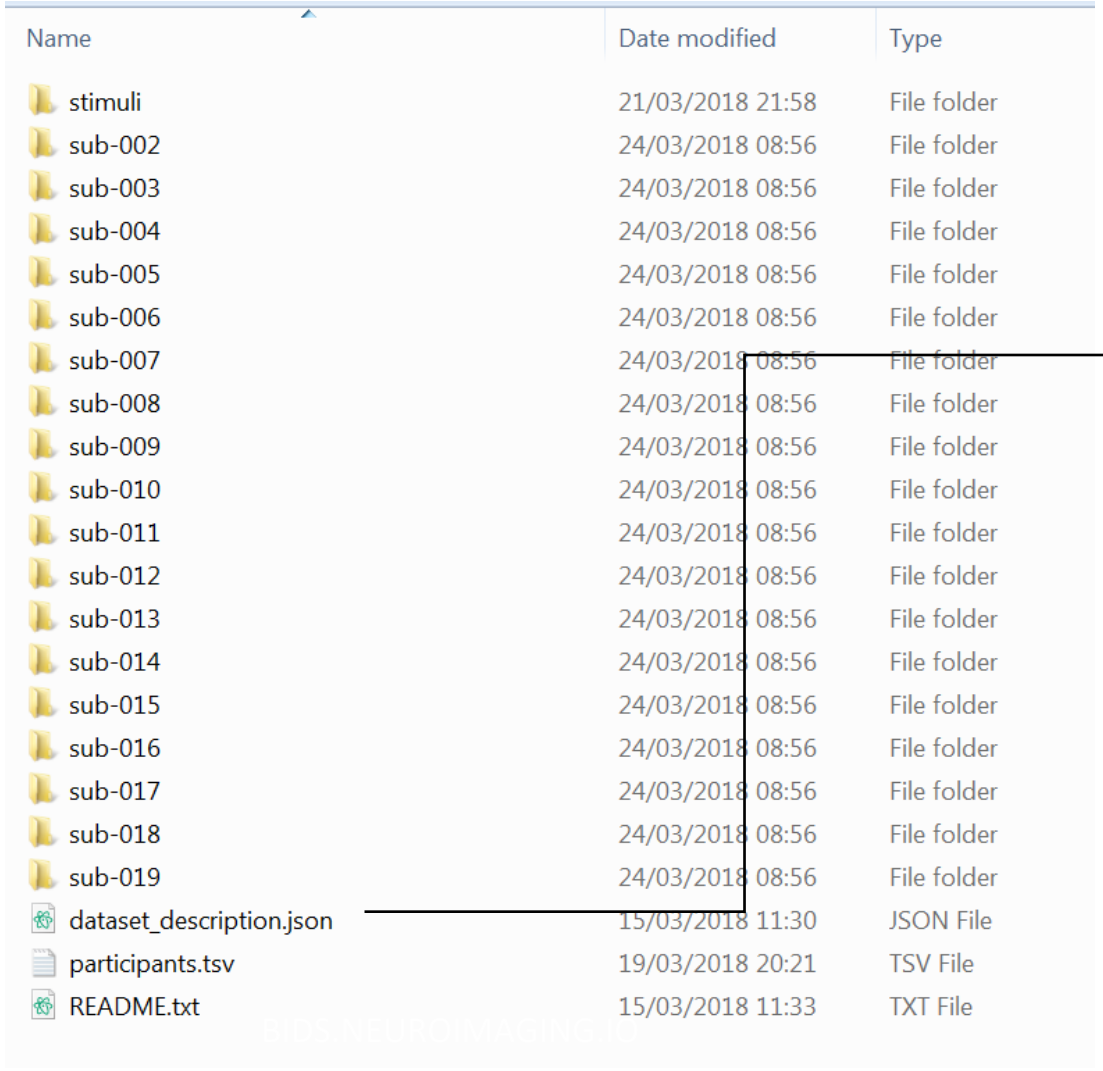


Name	Date modified	Type
stimuli	21/03/2018 21:58	File folder
sub-002	24/03/2018 08:56	File folder
sub-003	24/03/2018 08:56	File folder
sub-004	24/03/2018 08:56	File folder
sub-005	24/03/2018 08:56	File folder
sub-006	24/03/2018 08:56	File folder
sub-007	24/03/2018 08:56	File folder
sub-008	24/03/2018 08:56	File folder
sub-009	24/03/2018 08:56	File folder
sub-010	24/03/2018 08:56	File folder
sub-011	24/03/2018 08:56	File folder
sub-012	24/03/2018 08:56	File folder
sub-013	24/03/2018 08:56	File folder
sub-014	24/03/2018 08:56	File folder
sub-015	24/03/2018 08:56	File folder
sub-016	24/03/2018 08:56	File folder
sub-017	24/03/2018 08:56	File folder
sub-018	24/03/2018 08:56	File folder
sub-019	24/03/2018 08:56	File folder
dataset_description.json	15/03/2018 11:30	JSON File
participants.tsv	19/03/2018 20:21	TSV File
README.txt	15/03/2018 11:33	TXT File

- **source** (optional)
- **stimuli** (optional)
- **derivatives** (optional)
- **sub-XXX**

Anyone can now find his/her way around data !

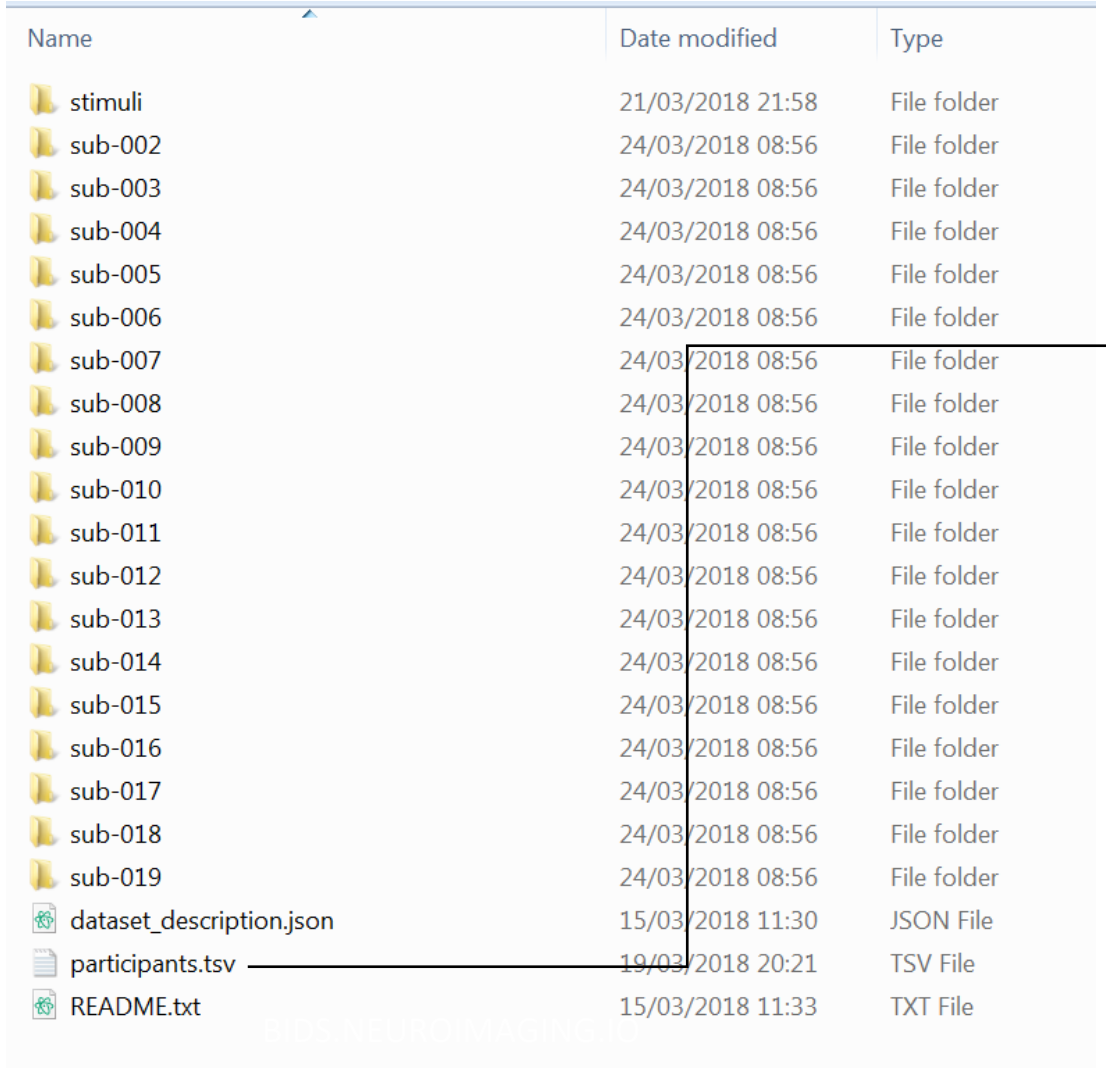
Metadata as text files (tsv, json) with standard dictionary




```
{
    "Name": "",
    "BIDSVersion": "",
    "License": "",
    "Authors": "",
    "Acknowledgements": "",
    "HowToAcknowledge": " ",
    "Funding": "",
    "ReferencesAndLinks": "",
    "SourceDatasetsURLs": ""
}
```

Your data are identifiable and citable

Metadata as text files (tsv, json) with standard dictionary

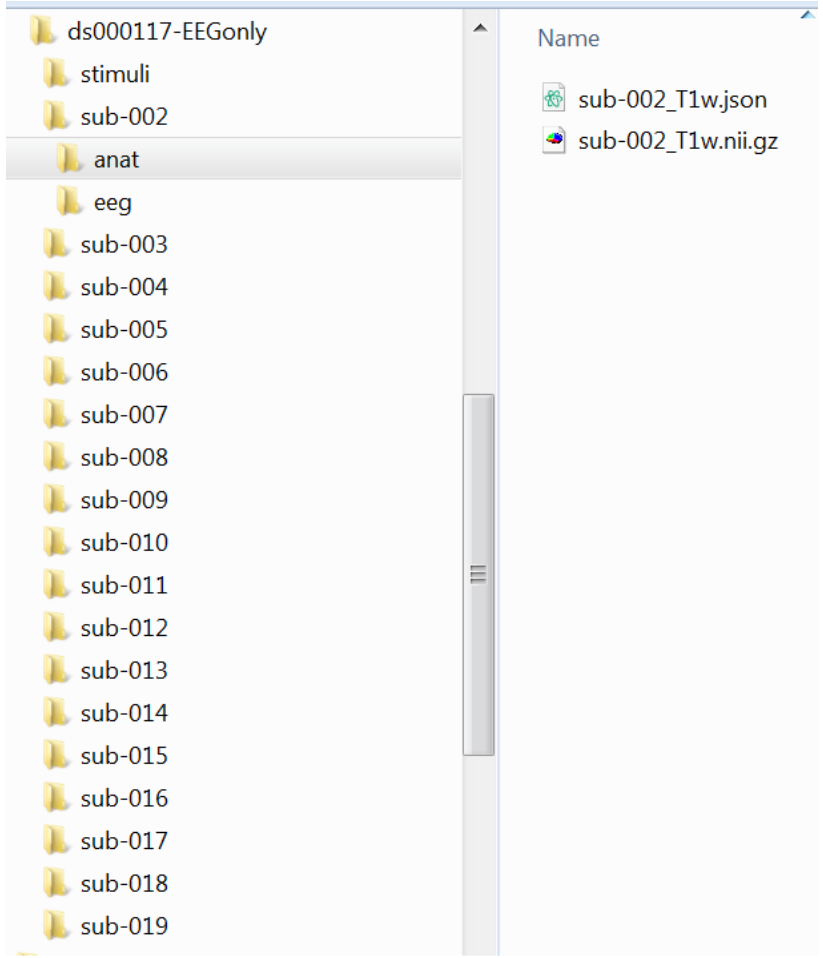



```
participant_id age sex
sub-002        34  M
sub-003        12  F
sub-004        33  F
```

Subjects info are shared at the root – easy to figure populations, age, and other basic demographics

Human and machine readable!

A BIDS folder (any/all modality)



Name redundancy – harder to make mistakes ; files comes with metadata (json)

Inside folder sub-002

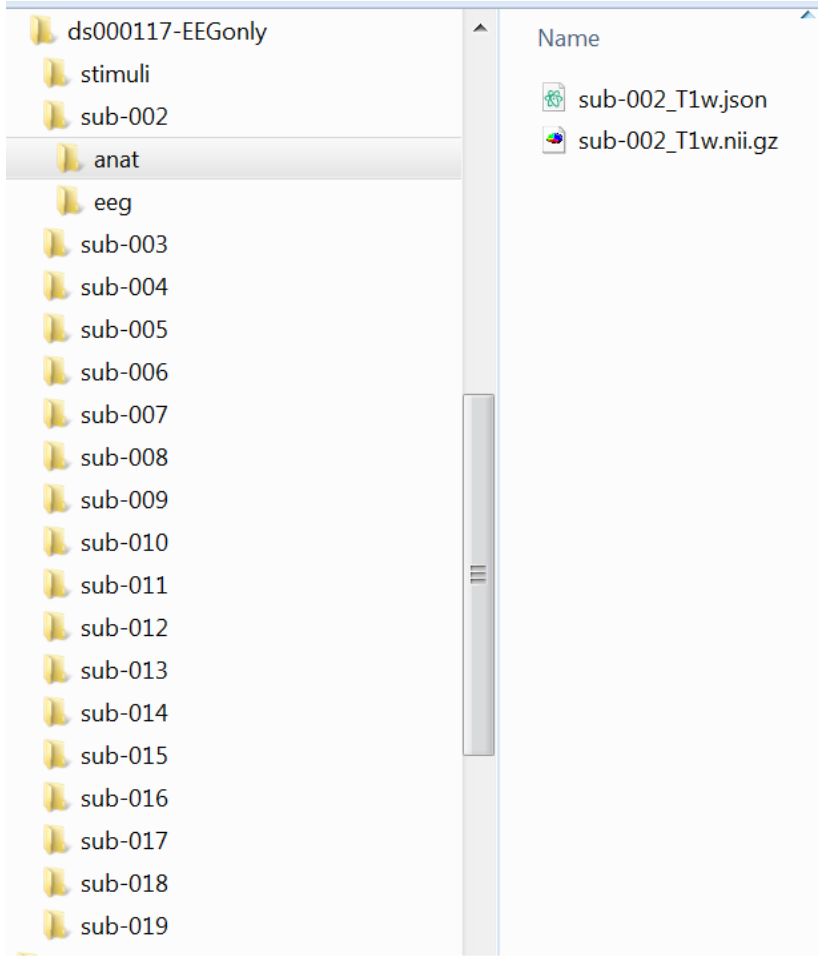
→ anat folder

sub-002_T1w.nii.gz

→ eeg folder

????

A BIDS folder (any/all modality)



Name redundancy – harder to make mistakes ; files comes with metadata (json)

Inside folder sub-002

→ anat folder

sub-002_T1w.nii.gz

→ eeg folder

sub-002_something_eeg.set

BIDS for EEG

www.nature.com/scientificdata

SCIENTIFIC DATA

OPEN

COMMENT

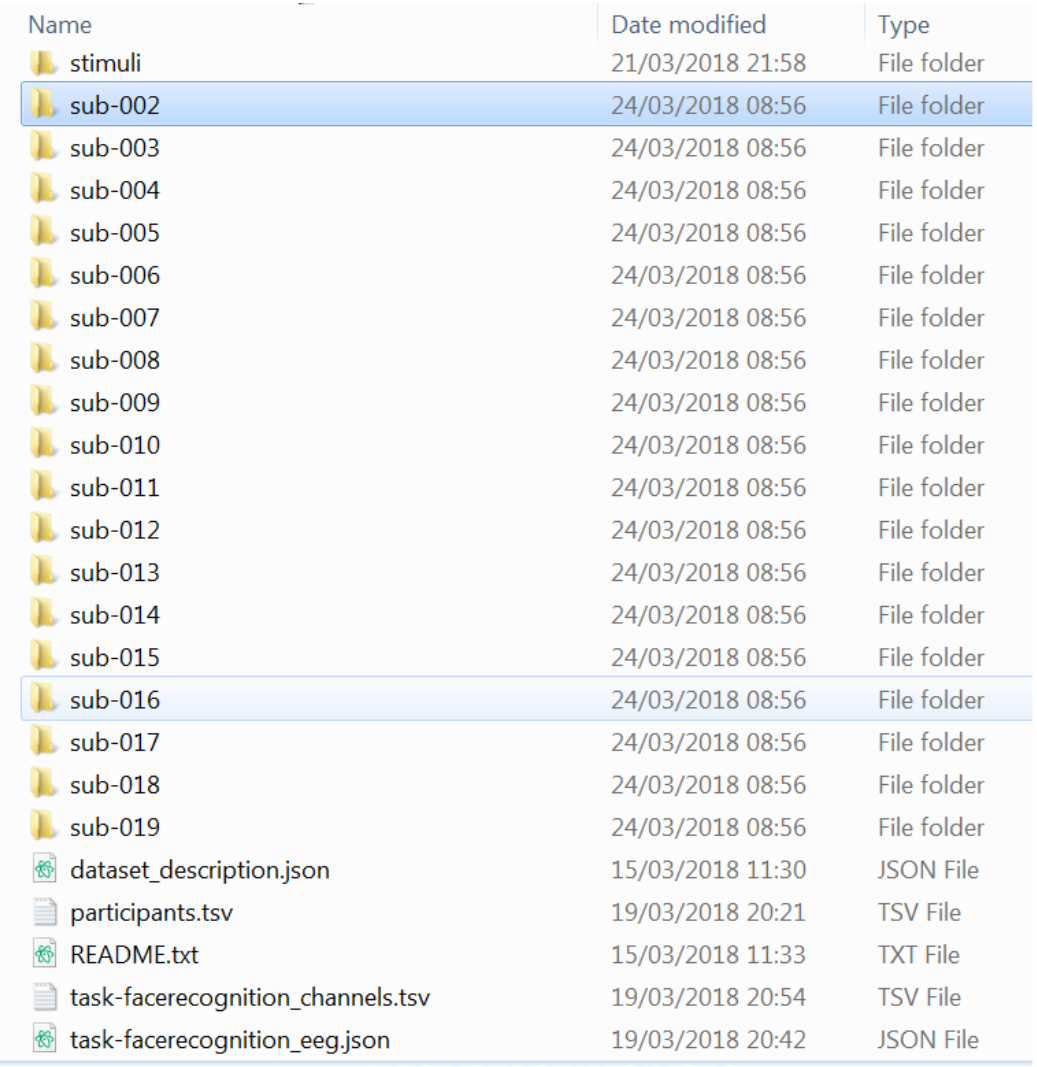
Received: 16 January 2019
Accepted: 7 May 2019
Published online: 25 June 2019

EEG-BIDS, an extension to the brain imaging data structure for electroencephalography

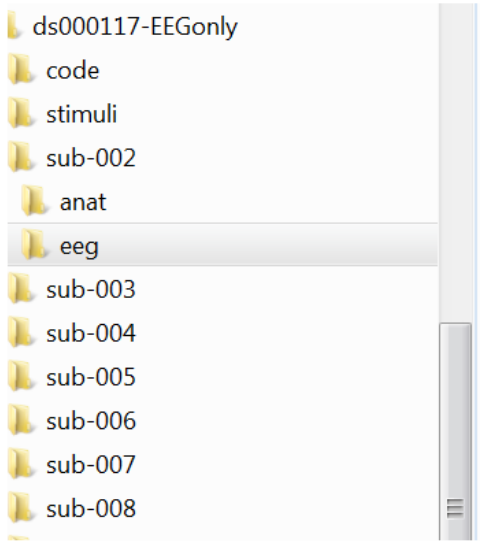
Cyril R. Pernet[1], Stefan Appelhoff[2], Krzysztof J. Gorgolewski[3], Guillaume Flandin[4], Christophe Phillips[5], Arnaud Delorme[6,7] & Robert Oostenveld[8,9]

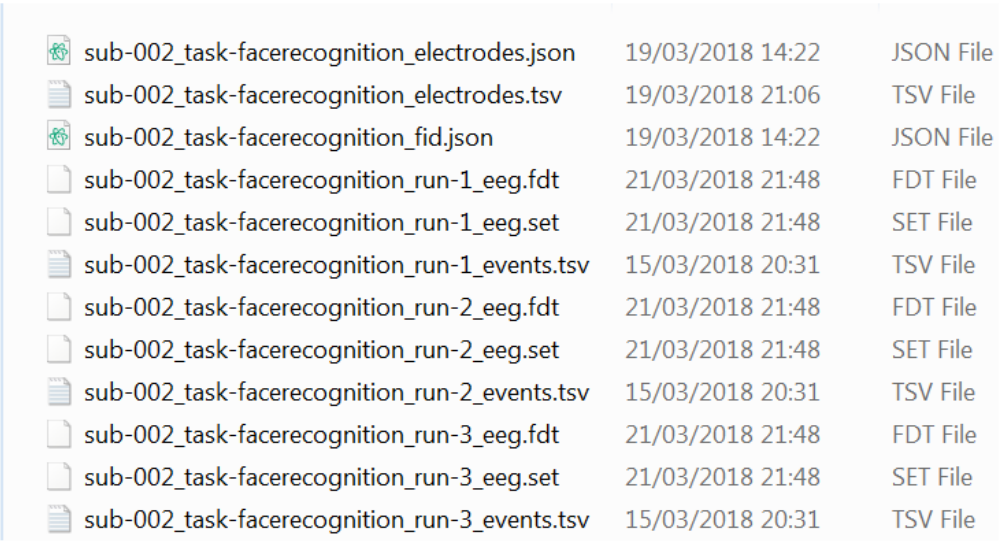
The Brain Imaging Data Structure (BIDS) project is a rapidly evolving effort in the human brain imaging research community to create standards allowing researchers to readily organize and share study data within and between laboratories. Here we present an extension to BIDS for electroencephalography (EEG) data, EEG-BIDS, along with tools and references to a series of public EEG datasets organized using this new standard.

EEG File formats (known and supported)



Name	Date modified	Type
stimuli	21/03/2018 21:58	File folder
sub-002	24/03/2018 08:56	File folder
sub-003	24/03/2018 08:56	File folder
sub-004	24/03/2018 08:56	File folder
sub-005	24/03/2018 08:56	File folder
sub-006	24/03/2018 08:56	File folder
sub-007	24/03/2018 08:56	File folder
sub-008	24/03/2018 08:56	File folder
sub-009	24/03/2018 08:56	File folder
sub-010	24/03/2018 08:56	File folder
sub-011	24/03/2018 08:56	File folder
sub-012	24/03/2018 08:56	File folder
sub-013	24/03/2018 08:56	File folder
sub-014	24/03/2018 08:56	File folder
sub-015	24/03/2018 08:56	File folder
sub-016	24/03/2018 08:56	File folder
sub-017	24/03/2018 08:56	File folder
sub-018	24/03/2018 08:56	File folder
sub-019	24/03/2018 08:56	File folder
dataset_description.json	15/03/2018 11:30	JSON File
participants.tsv	19/03/2018 20:21	TSV File
README.txt	15/03/2018 11:33	TXT File
task-facerecognition_channels.tsv	19/03/2018 20:54	TSV File
task-facerecognition_eeg.json	19/03/2018 20:42	JSON File





sub-002_task-facerecognition_electrodes.json	19/03/2018 14:22	JSON File
sub-002_task-facerecognition_electrodes.tsv	19/03/2018 21:06	TSV File
sub-002_task-facerecognition_fid.json	19/03/2018 14:22	JSON File
sub-002_task-facerecognition_run-1_eeg.fdt	21/03/2018 21:48	FDT File
sub-002_task-facerecognition_run-1_eeg.set	21/03/2018 21:48	SET File
sub-002_task-facerecognition_run-1_events.tsv	15/03/2018 20:31	TSV File
sub-002_task-facerecognition_run-2_eeg.fdt	21/03/2018 21:48	FDT File
sub-002_task-facerecognition_run-2_eeg.set	21/03/2018 21:48	SET File
sub-002_task-facerecognition_run-2_events.tsv	15/03/2018 20:31	TSV File
sub-002_task-facerecognition_run-3_eeg.fdt	21/03/2018 21:48	FDT File
sub-002_task-facerecognition_run-3_eeg.set	21/03/2018 21:48	SET File
sub-002_task-facerecognition_run-3_events.tsv	15/03/2018 20:31	TSV File

Anyone can read the data

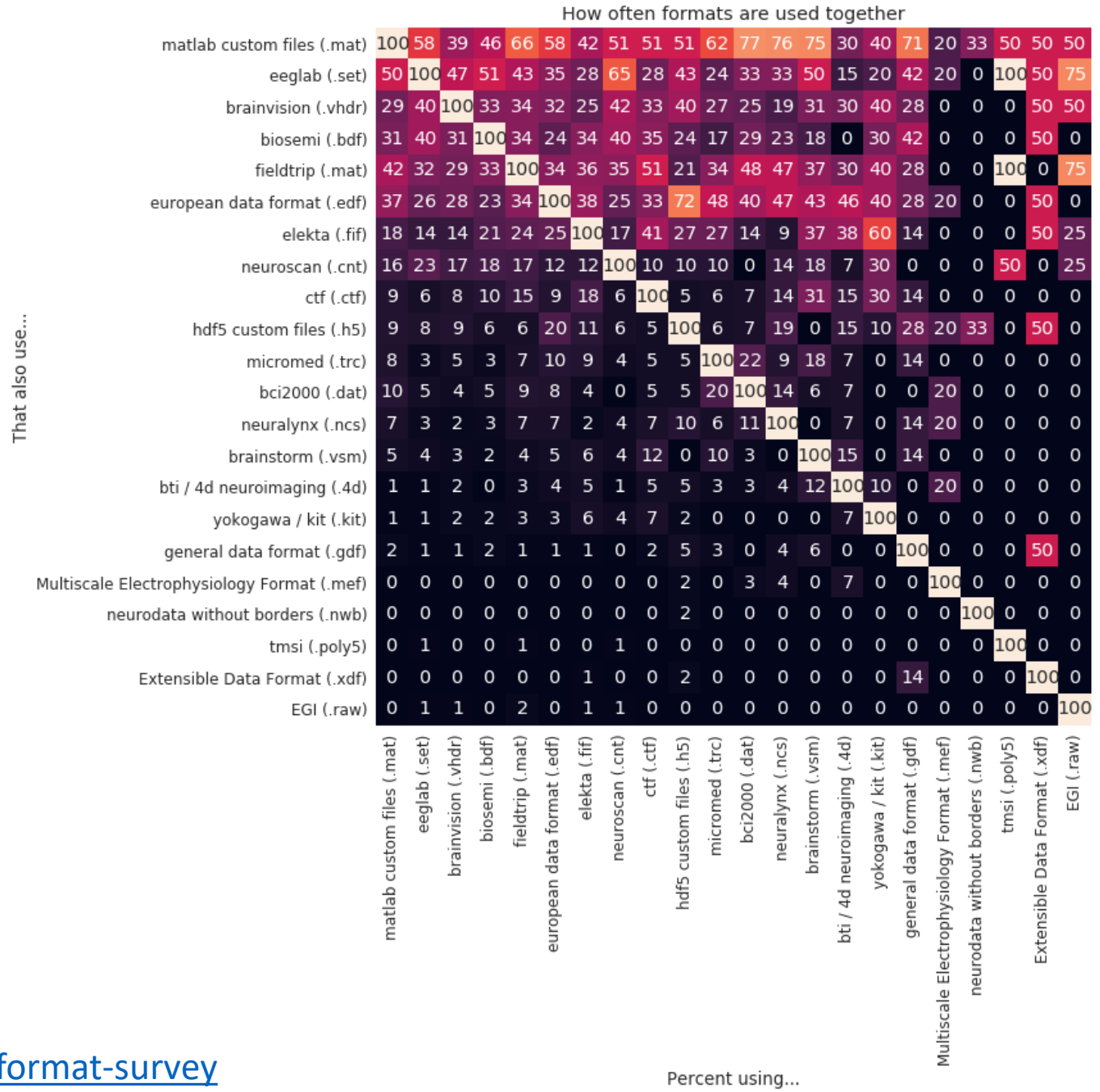
- **official: edf & vhdr/eeg**
- **unofficial: set/fdt, bdf**
  (all supported by open-source software)

(i)EEG File formats

(i) wide usage in the community

(ii) open access documentation, open source implementation for both reading and writing in at least two programming languages and widely supported in multiple software packages (both open source and commercial)

(iii) high numerical precision (16 and 32 bits respectively).

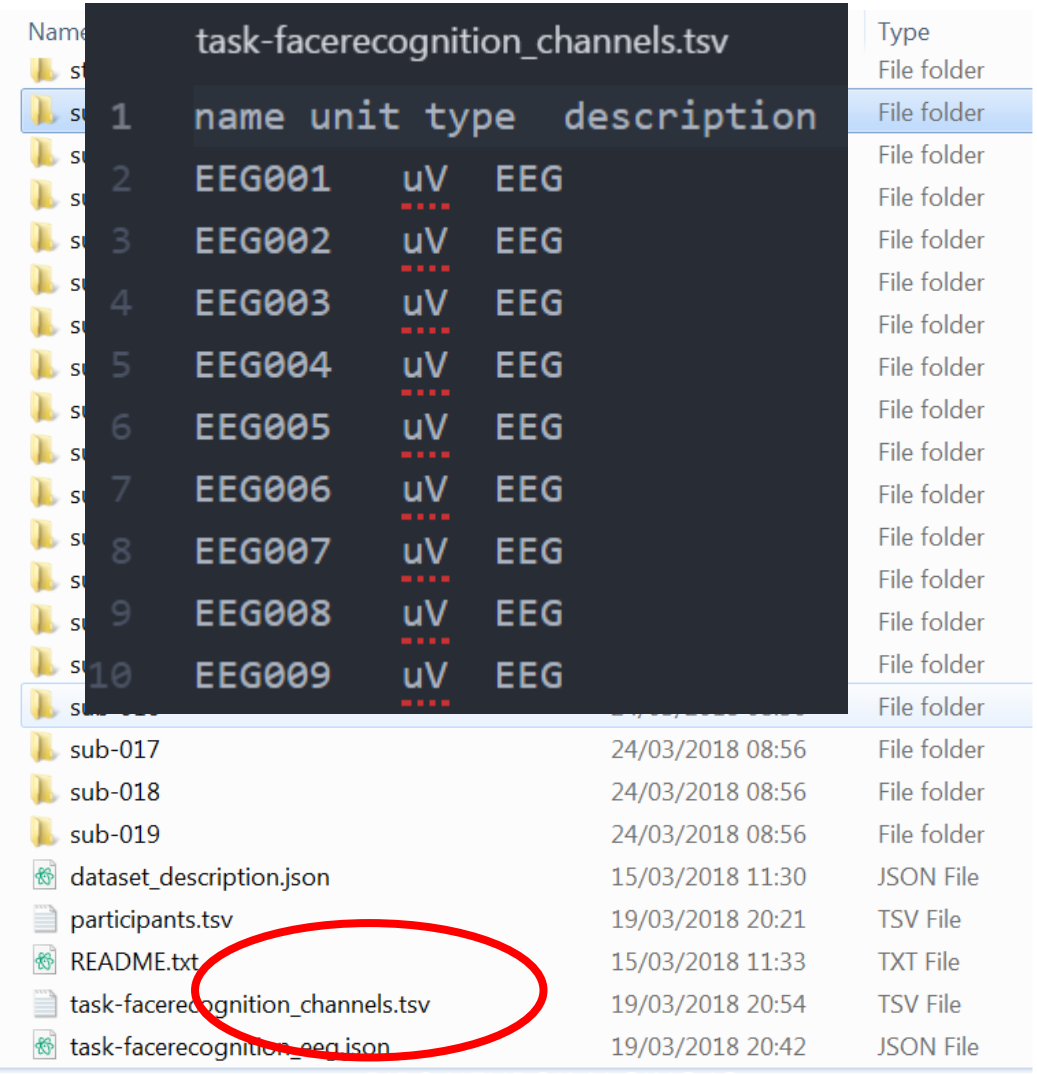


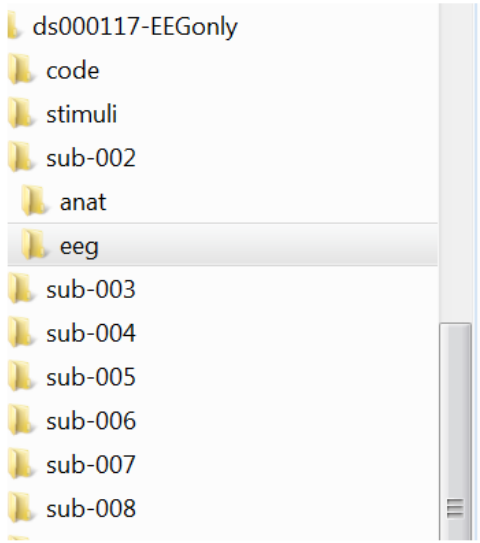
https://bids.berkeley.edu/news/bids-megeegieeg-data-format-survey

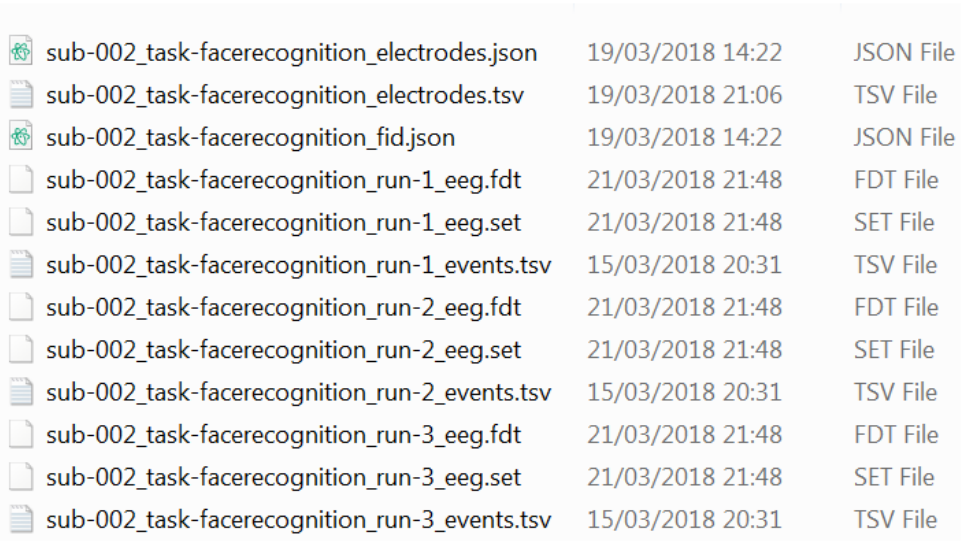
Some (M/iE/E)EG specifics metadata

task-facerecognition_channels.tsv

	name	unit	type	description
2	EEG001	uV	EEG	
3	EEG002	uV	EEG	
4	EEG003	uV	EEG	
5	EEG004	uV	EEG	
6	EEG005	uV	EEG	
7	EEG006	uV	EEG	
8	EEG007	uV	EEG	
9	EEG008	uV	EEG	
10	EEG009	uV	EEG	







Some (M/iE/E)EG specifics metadata

task-facerecognition_channels.tsv

	name	unit	type	description
2	EEG001	uV	EEG	
3	EEG002	uV	EEG	
4	EEG003	uV	EEG	
5	EEG004	uV	EEG	
6	EEG005	uV	EEG	
7	EEG006	uV	EEG	
8	EEG007	uV	EEG	
9	EEG008	uV	EEG	
10	EEG009	uV	EEG	

sub-002_task-facerecognition_electrodes.tsv

	name	x	y	z
2	EEG001	-0.0416675	-0.0661554	0.0167599
3	EEG002	0.00464		
4	EEG003	0.0419		
5	EEG004	-0.046		
6	EEG005	-0.025		
7	EEG006	0.0067		
8	EEG007	0.0386		
9	EEG008	0.0564		
10	EEG009	-0.057		

sub-002_task-facerecognition_fid.json

```
{
    "EEGCoordinateSystem": "T1w",
    "EEGCoordinateUnits": "mm",
    "AnatomicalMRICoordinateSystem": "T1w",
    "AnatomicalMRICoordinateUnits": "mm",
    "LandmarkCoordinates": "{\"LPA\":[-0.070999 -2.6537e-08 -
    "LandmarkCoordinateSystem": "T1w",
    "LandmarkCoordinateUnits": "mm"
}
```

Name | Type

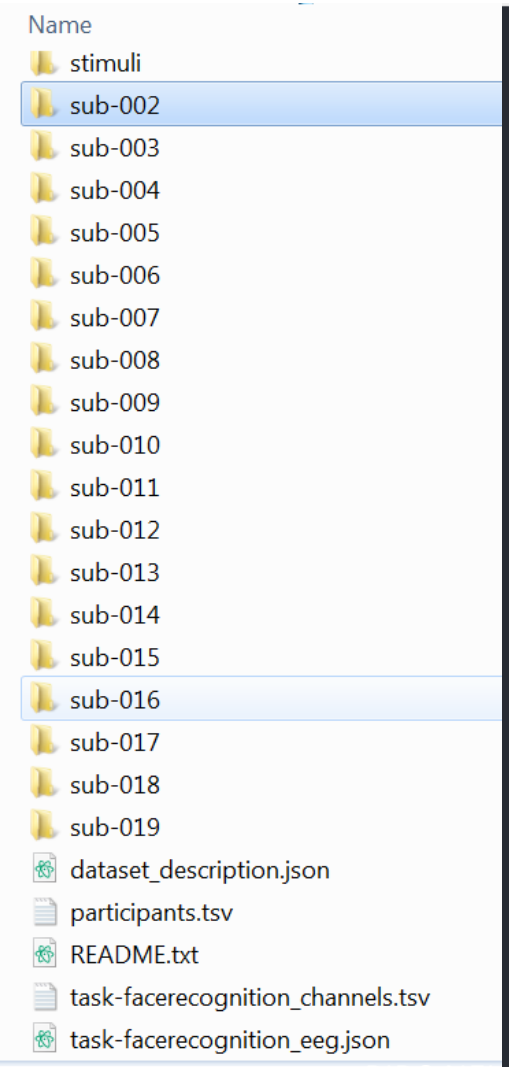
- sub-017 — File folder
- sub-018 — File folder
- sub-019 — File folder
- dataset_description.json
- participants.tsv — 19/03/2018 20:21 — TSV File
- README.txt — 15/03/2018 11:33 — TXT File
- task-facerecognition_channels.tsv — 19/03/2018 20:54 — TSV File
- task-facerecognition_eeg.json — 19/03/2018 20:42 — JSON File

ds000117-EEGonly

code

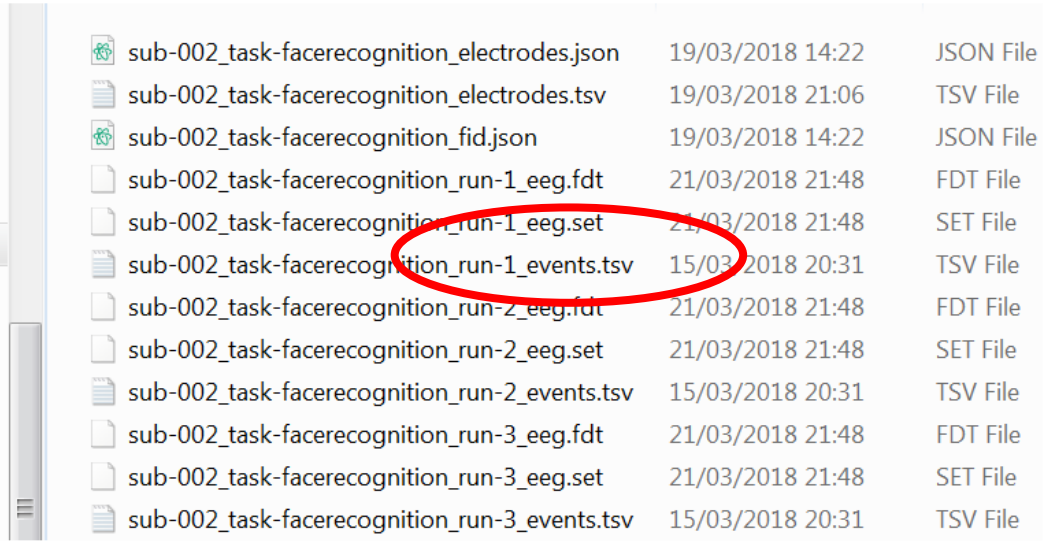
- sub-002_task-facerecognition_electrodes.json — 19/03/2018 14:22 — JSON File
- sub-002_task-facerecognition_electrodes.tsv — 19/03/2018 21:06 — TSV File
- k-facerecognition_fid.json — 19/03/2018 14:22 — JSON File
- k-facerecognition_run-1_eeg.fdt — 21/03/2018 21:48 — FDT File
- k-facerecognition_run-1_eeg.set — 21/03/2018 21:48 — SET File
- k-facerecognition_run-1_events.tsv — 15/03/2018 20:31 — TSV File
- k-facerecognition_run-2_eeg.fdt — 21/03/2018 21:48 — FDT File

Some (M/iE/E)EG specifics metadata



sub-002_task-facerecognition_run-1_events.tsv

	onset	trial_type	eventcode
1	26624	unfamiliar_new	13
2	27667	response	0
3	29968	unfamiliar_second_early	14
4	30680	response	0
5	33386	unfamiliar_new	13
6	36694	unfamiliar_new	13
7	40205	famous_new	5
8	41041	response	0
9	43530	unfamiliar_new	13
10	47076	famous_new	5
11	47897	response	0
12	50641	scrambled_new	17
13	54041	unfamiliar_second_late	15
14	57624	famous_new	5
15	58404	response	0
16	61097	unfamiliar_second_late	15
17	64680	unfamiliar_new	13
18	65488	response	0
19	68098	unfamiliar_second_early	14



Yes, this info is often present on a channel

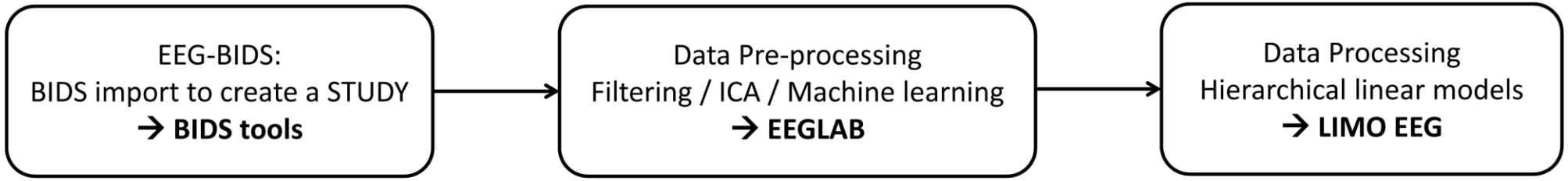
- **as triggers, no metadata/name**
- **also easier to figure out as user (i.e. yourself in 6 months) what those event codes are**

Prof. Smith (2030)

What other benefits?

Benefits

- Use each other data / well documented / ready to analyse
- You can have **pipelines!**



- Workflow: the sequence of computational steps through which a piece of work passes from initiation to completion.
- Pipeline: set of data processing elements connected in series, where the output of one element is the input of the next one – pipelines implement workflows.

EEGLAB-LIMO pipeline

TECHNOLOGY AND CODE ARTICLE

Front. Neurosci. | doi: 10.3389/fnins.2020.610388

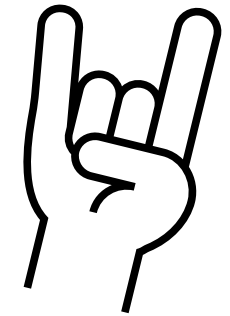
From BIDS-formatted EEG data to sensor-space group results: a fully reproducible workflow with EEGLAB and LIMO EEG

Provisionally accepted

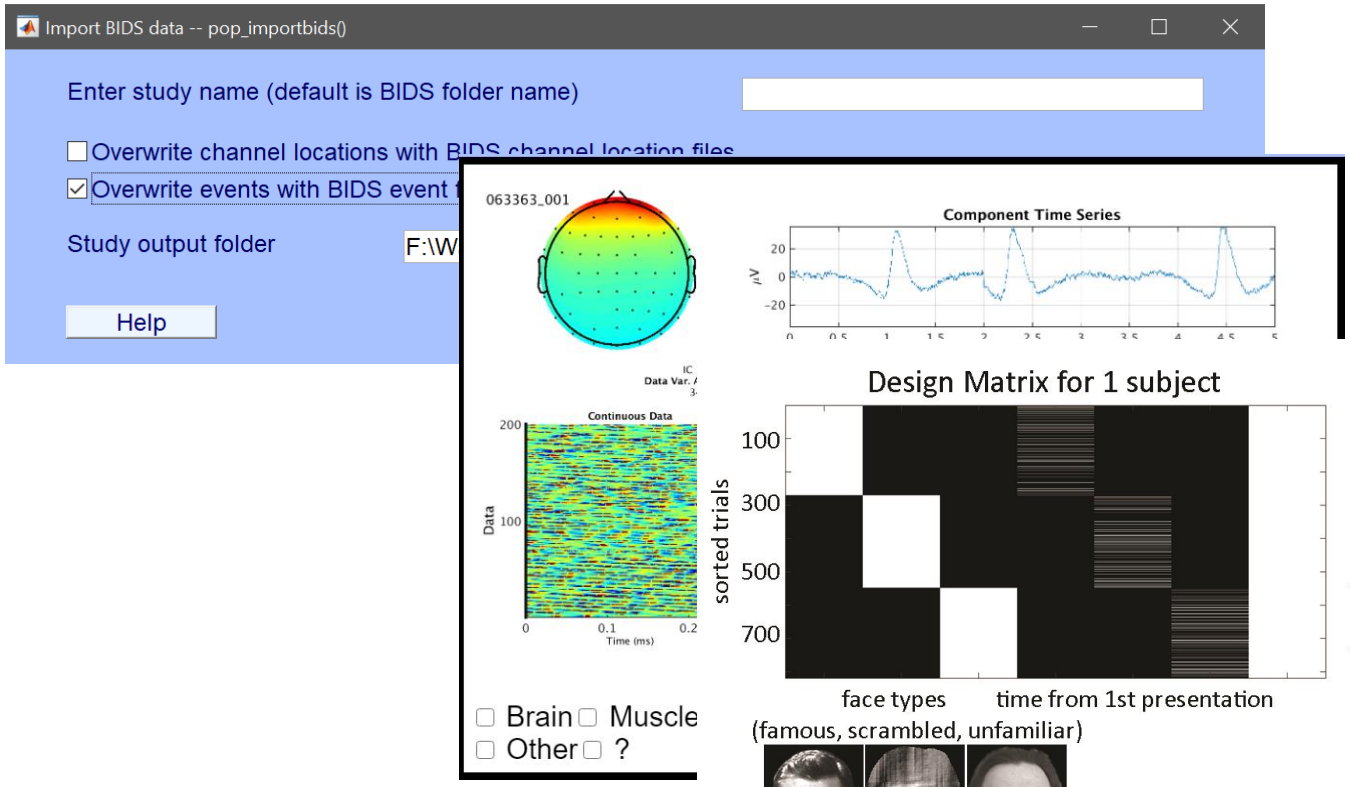
 Cyril R. Pernet[1*],  Ramón Martínez-Cancino[2], Dung Truong[2],  Scott Makeig[2] and  Arnaud Delorme[2]

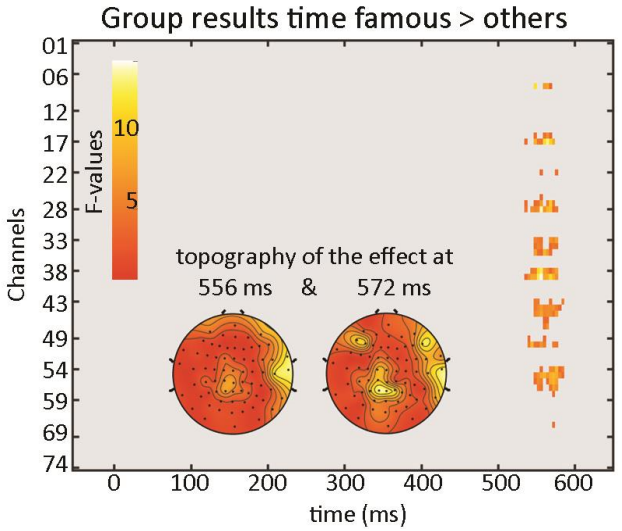
EEGLAB-LIMO pipeline

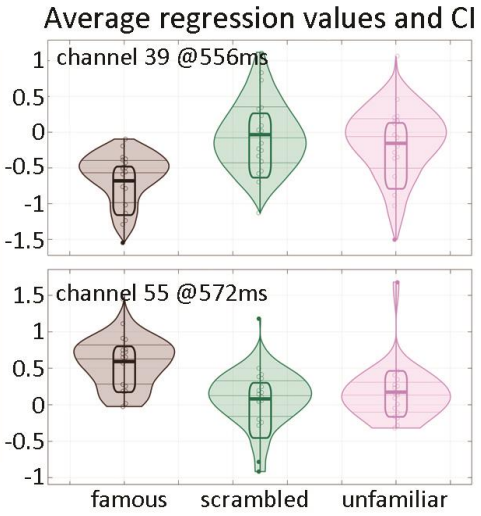
Fully reproducible code from import BIDS to figures



Attend our session this Thursday to check this out

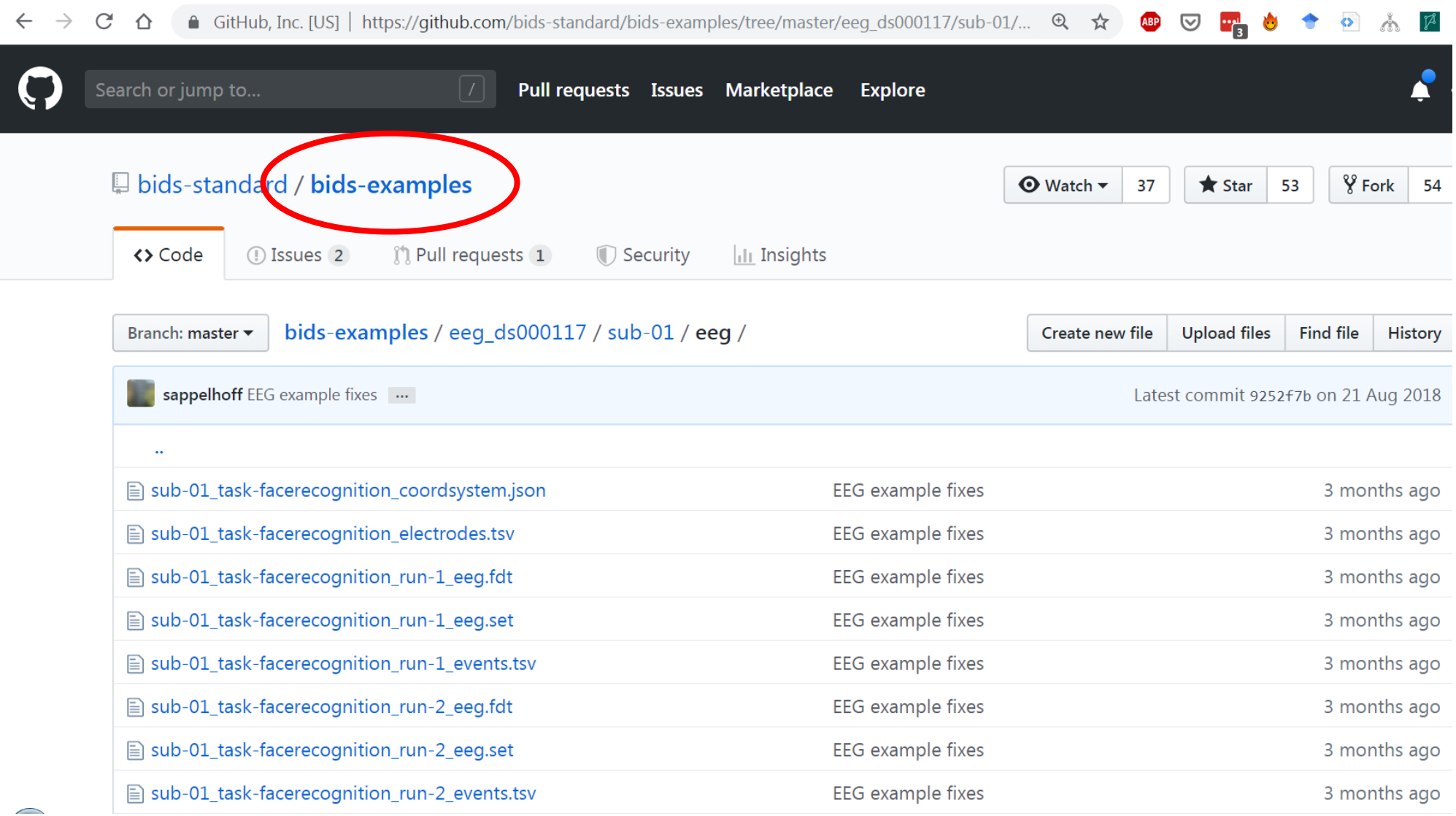




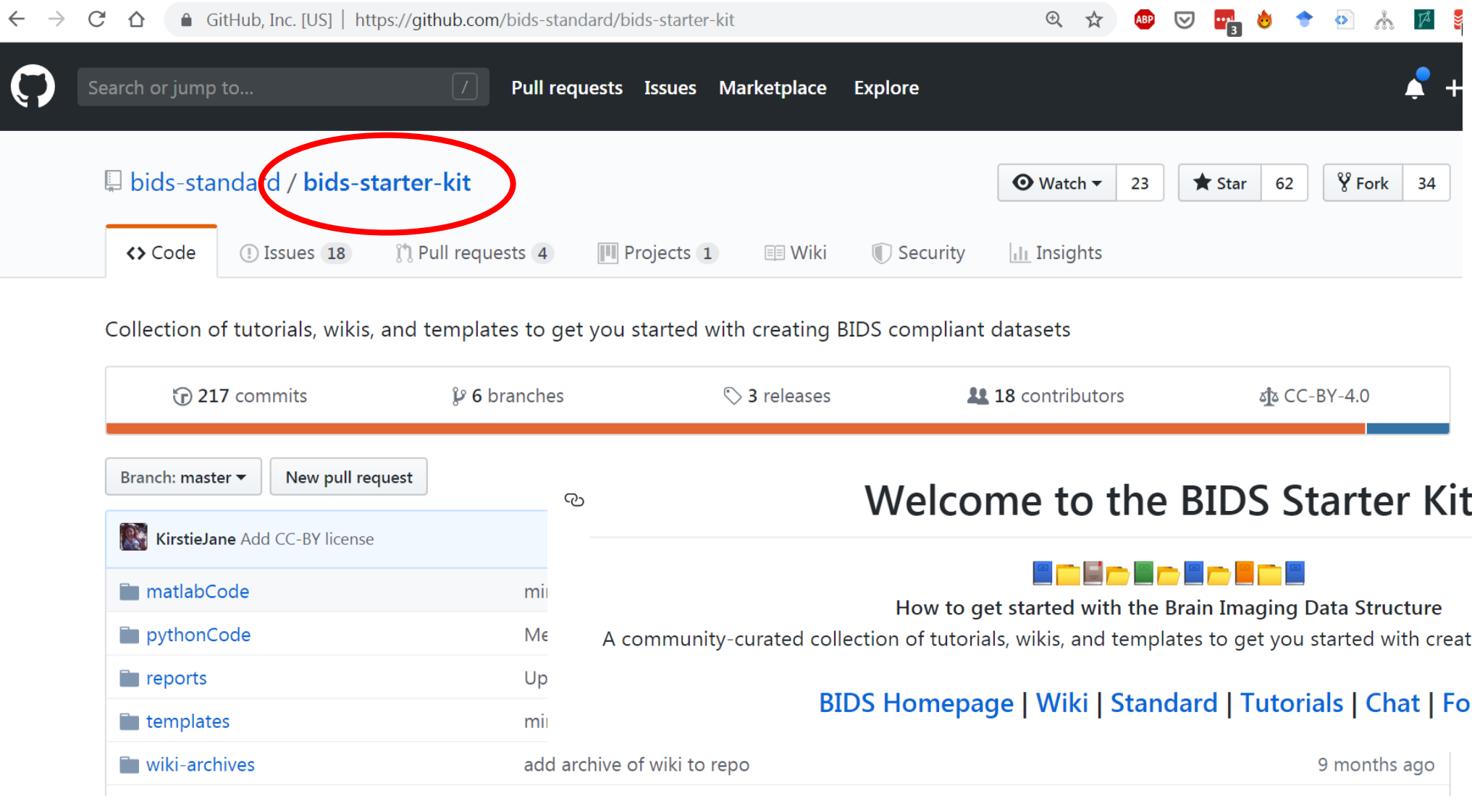


How do I start?

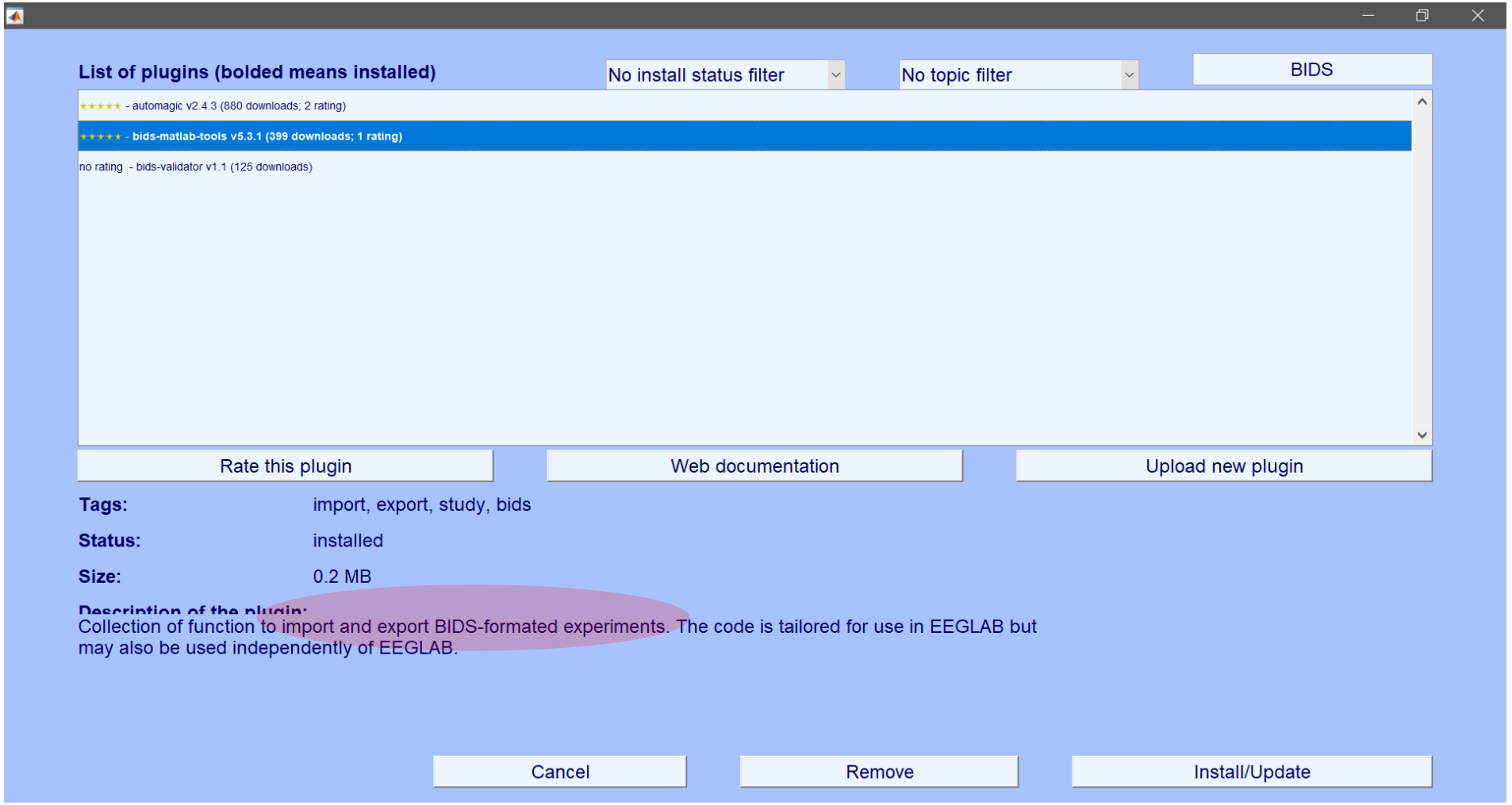
Preparing data using BIDS



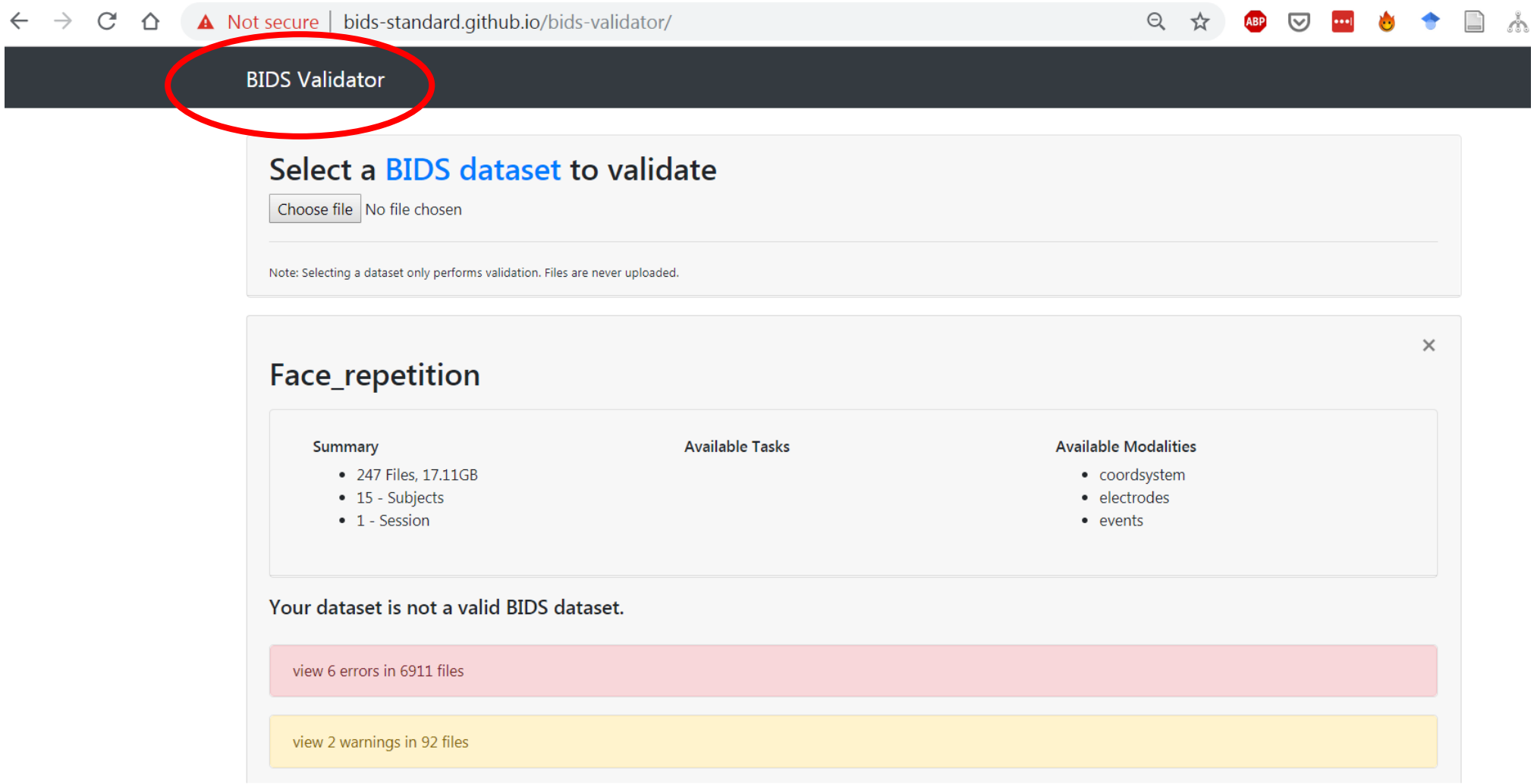
Preparing data using BIDS



Preparing data using BIDS



Preparing data using BIDS



Sharing data using BIDS

Findable

Data and supplementary materials have sufficiently rich metadata and a unique and persistent identifier.

Accessible

Data is deposited in a trusted repository.

Interoperable

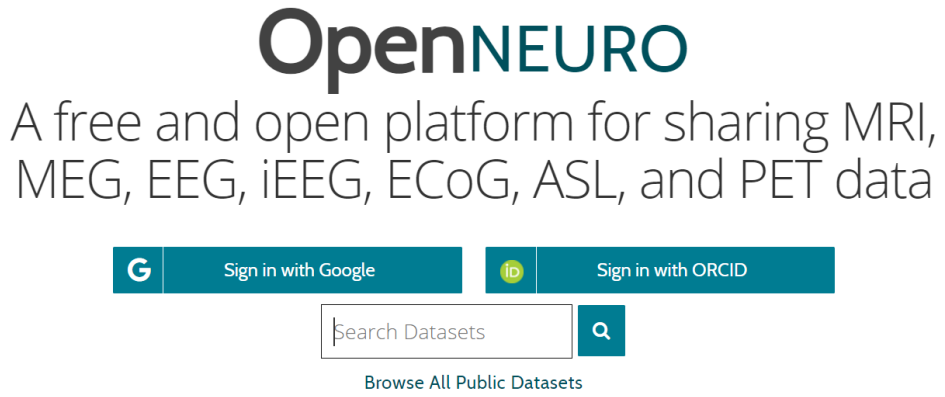
(Meta)data uses a formal, shared, and broadly applicable language or format.

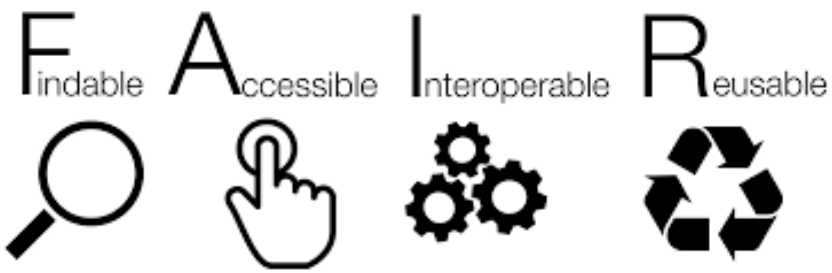
Reusable

Data is described with clear and understandable attributes.

There should be a clear and acceptable license for re-use.

}





https://www.force11.org/group/fairgroup/fairprinciples

Delorme – 2020 BIDS talk